Compressive Oversampling for Robust Data Transmission in Sensor Networks

Zainul Charbiwala, Supriyo Chakraborty, Sadaf Zahedi, Younghun Kim, Mani B. Srivastava
University of California, Los Angeles

Ting He, Chatschik Bisdikian
IBM T. J. Watson Research Center

Abstract—**Data loss in wireless sensing applications is inevitable and while there have been many attempts at coping with this issue, recent developments in the area of Compressive Sensing (CS) provide a new and attractive perspective. Since many physical signals of interest are known to be sparse or compressible, employing CS, not only compresses the data and reduces effective transmission rate, but also improves the robustness of the system to channel erasures. This is possible because reconstruction algorithms for compressively sampled signals are not hampered by the stochastic nature of wireless link disturbances, which has traditionally plagued attempts at proactively handling the effects of these errors. In this paper, we propose that if CS is employed for source compression, then CS can further be exploited as an application layer erasure coding strategy for recovering missing data. We show that CS erasure encoding (CSEC) with random sampling is efficient for handling missing data in erasure channels, paralleling the performance of BCH codes, with the added benefit of graceful degradation of the reconstruction error even when the amount of missing data far exceeds the designed redundancy. Further, since CSEC is equivalent to nominal oversampling in the incoherent measurement basis, it is computationally cheaper than conventional erasure coding. We support our proposal through extensive performance studies.**

Keywords-erasure coding; compressive sensing.

I. Introduction

Data loss in wireless sensing applications is inevitable, either due to exogenous (such as transmission medium impediments) or endogenous (such as faulty sensors) causes. While many schemes have been proposed to cope with this issue, the emerging area of compressive sensing enables a fresh perspective for sensor networks. Many physical phenomena are compressible in a known domain and it is beneficial to use some form of source coding or compression, whenever practical, to reduce redundancy in the data prior to transmission. For example, sounds are compactly represented in the frequency domain whereas images may be compressed in the wavelet domain. Traditionally, compression is performed at the application layer after the signal is sampled and digitized and typically imposes a high computation overhead at the encoder. This cost is the major reason that low-power embedded sensing systems have to make a judicious choice about when to employ source coding [14]. Advances in compressive sensing (CS) [5], however, have made it possible to shift this computation burden to the decoder, presumably a more capable data sink (e.g., a wireless sensor network's base station), which is neither power nor memory bound. CS enables source compression to be performed inexpensively at the encoder, with a slight sampling overhead[1] and with little or no knowledge of the compression domain.

Compression, however, also makes each transmitted bit of information more precious, necessitating a reliable transport mechanism to maintain the quality of information. To cope with channel disturbances, retransmission schemes have popularly been applied, but they are inefficient in many scenarios, such as on acoustic links used for underwater communication [1], where round trip delays and ARQ traffic cost precious throughput. Retransmissions are ineffective in other cases too, for example, in multicast transmissions or when transmission latency is paramount for rapid detection. Forward error correction schemes like Reed-Solomon [15], LT [12] or convolutional codes are better suited for these scenarios, but their use in low-power sensing has been limited, primarily because of their computational complexity or bandwidth overhead [16].

Fortunately, the computational benefits of CS coupled with its inherent use of randomness can make it an attractive choice for combating erasures as well. A key observation that makes this possible is that reconstruction algorithms for compressively sampled data exploit randomness within the measurement process. Therefore, the stochastic nature of wireless link losses and short-term sensor malfunctions do not hamper the performance of reconstruction algorithms at the decoder. In fact, to the decoder, losses are indistinguishable from an *a priori* lower sensing rate. We, therefore, propose using compressive sensing as a low encoding-cost, proactive erasure coding strategy and show, in particular, that employing CS erasure coding (CSEC) has three desirable features:

- CSEC is achieved by nominal oversampling in an incoherent measurement basis. Compared to the cost of conventional erasure coding that is applied over the entire data set from scratch, additional sampling can be much cheaper, especially if random sampling is used. The high cost of CS decoding is amortized over joint source and channel coding and is free, if CS was already being employed for source decompression.

This material is supported in part by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001, the U.S. Office of Naval Research under MURI-VA Tech Award CR-19097-430345, the National Science Foundation under grant CCF-0820061, and the UCLA Center for Embedded Networked Sensing. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the listed funding agencies. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes not withstanding any copyright notation herein.

[1] CS sampling incurs a logarithmic overhead when compared to acquiring the signal directly in the compression domain.

Report Documentation Page				*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.				
1. REPORT DATE **2010**		2. REPORT TYPE		3. DATES COVERED **00-00-2010 to 00-00-2010**
4. TITLE AND SUBTITLE **Compressive Oversampling for Robust Data Transmission in Sensor Networks**				5a. CONTRACT NUMBER
				5b. GRANT NUMBER
				5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)				5d. PROJECT NUMBER
				5e. TASK NUMBER
				5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California, Los Angeles,Los Angeles,CA,90095**				8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**				
13. SUPPLEMENTARY NOTES **The 29th Conference on Computer Communications (INFOCOM) , March 2010. (TR-UCLA-NESL-200912-01)**				
14. ABSTRACT **see report**				
15. SUBJECT TERMS				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**9**	

- The performance of CS erasure coding with random sampling is similar to conventional schemes such as Reed-Solomon and, in general, the BCH family of codes, in that it can recover as many missing symbols for the same relative redundancy in a memoryless erasure channel. This aspect is covered in Sec. II.D.
- CS erasure coding is robust to estimation error in channel loss probability. For example, if a BCH code of block size n were designed to correct up to t erasures, in a situation where $e > t$ erasures occur, the entire block of n symbols would be discarded. This implies that BCH codes must consider and be designed for a worst-case loss probability for recovery to succeed. An equivalent CS strategy, however, guarantees that even if $e > t$ symbols are lost, the best approximation of the signal is reconstructed from the remaining $n-e$ symbols. This means that even if channel coding fails at the physical layer, CSEC can recover the signal at the application layer.

Despite its advantages, CS erasure coding is not intended as a replacement for traditional physical layer channel codes. It is neither as general-purpose (i.e. it cannot be used for arbitrary non-sparse data), nor is the decoding as computationally efficient (yet). Instead, CSEC should be considered as a coding strategy that is applied at the application layer, where it utilizes knowledge of signal characteristics for better performance. In this regard, it is the reduced encoding cost that makes CSEC especially attractive for low-power embedded sensing. We quantify its energy efficiency benefits in Sec. III.C.

We highlight the conventional and proposed approaches in Figs. 2 and 1 respectively. Notation used in the figures is introduced in the Sec. II. Typically, source coding is performed after the signal is completely acquired, removing redundancy in the samples through a lossy or lossless compression routine. This step is performed at the application layer and utilizes known signal models to determine the most succinct representation domain for the phenomenon of interest. The compressed data is then handed to the communication protocol stack, where just before transmission, usually at the physical layer, the data may be encoded again to introduce a controlled amount of redundancy. If transmitted symbols are received in error or not at all, the decoder may be able to recover the original data using this extra information.

If, on the other hand, compressive sampling were to be employed for joint source and channel coding, the sampling stage would itself subsume the coding blocks. CS sampling uses one of a variety of random measurement techniques that ensure that sufficient unique information is captured by the sampling process with high probability. We propose that the CS sampling block should be designed not merely to include prior knowledge of signal characteristics in terms of its sparsity in a specific domain, but consider channel characteristics as well. In particular, we propose tuning the sampling process, e.g., through judicious oversampling, to improve the robustness to channel impairments. We show in Sec. II.D that the universality of compressive sensing to the sparsity domain extends to the channel model as well, making CSEC advantageous even when channel characteristics are not precisely known. In particular, we will see in Sec. III.B that signal reconstruction performance with CSEC degrades gracefully when the average sampling rate at the acquisition stage is insufficient for exact recovery.

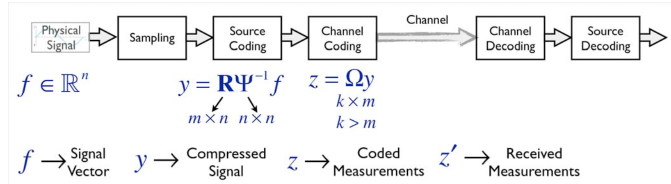

Figure 2. The conventional sequence of source and channel coding.

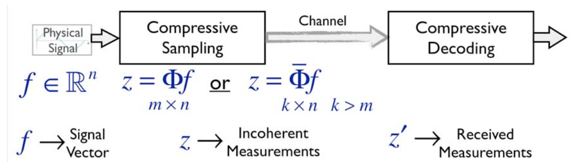

Figure 1. Proposed joint source-channel coding using compressive sensing.

II. Compressive Sensing for Erasure Coding

The problem we seek to address is acquiring a length n signal vector $f \in \mathbb{R}^n$ at a sensor node and communicating a length k measurement vector $z \in \mathbb{R}^k$ such that f can be recovered accurately at a base station one or more wireless hops away. We assume a generic wireless sensing application, where the signals are sparse or compressible in a known domain, and the data is collected centrally at a capable base-station. To construct our argument, we first briefly discuss both channel coding and compressive sensing. We then propose a compressive coding strategy in which oversampling suffices for robust data transmission.

A. Channel Coding Overview

If we consider a simple sense-and-send scenario where we send the sensed signal $f \in \mathbb{R}^n$ to a base station over an unreliable communication channel, $z = f$ and $k = n$. However, if a channel coding function $\mathbb{F}_c$ is applied prior to transmission, $z = \mathbb{F}_c(f)$ and since channel coding increases the average transmission rate by adding redundancy, $k \geq n$. Consider a linear channel coding function $z = \Omega f$, where $\Omega \in \mathbb{R}^{k \times n}$ is the equivalent channel coding matrix. When z is transmitted through a lossy channel, some measurements may not be received at the other end. We define the received measurement vector z' of length $k' = k - e$, where e is the number of erasures. The channel may also be modeled as a linear operator $\mathbf{C} \in \mathbb{R}^{k' \times k}$ so that $z' = \mathbf{C}z$. In general, $\mathbf{C}$ can consist of any values, but for the class of erasure channels we consider here, $\mathbf{C}$ is a submatrix of an identity matrix $\mathbf{I}_k$, where e rows have been omitted. Recovering the original signal from the received data is then a decoding operation of the form:

$$\hat{f} = \left(\mathbf{C}\Omega\right)^{+} z', \tag{1}$$

where, $X^{+} = \left(X^T X\right)^{-1} X^T$ is the Moore-Penrose pseudoinverse. If $\mathbf{C}\Omega$ is full rank, the decoding will be successful, else, the signal f cannot be recovered and the measurement vector z' is discarded. Based on the application, the encoder may either resend z or may re-encode f with a higher redundancy code before retransmitting.

We would like to emphasize a property of erasure channels and linear coding here. Data that is missing from vector z is caused by the channel matrix $\mathbf{C}$, which is generated by omitting rows from an identity matrix $\mathbf{I_k}$ at the indices corresponding to the missing data. But, since z is formed using $z = \Omega f$, one may instead view the combined coding and loss process as one of coding alone, where $\Omega' = \mathbf{C}\Omega$ is the equivalent coding matrix generated from Ω by omitting $k - k'$ of its rows at the indices corresponding to the lost data. This means that missing data at the receiver can be considered the same as not having those rows in Ω to begin with. We will use this perspective later when we discuss properties of compressive sensing.

Now, if we knew that the signal f contained redundancy, we could have compressed it before channel coding. We represent f using a sparse vector $x \in \mathbb{R}^n$, by transforming it through an ortho-normal basis $\Psi \in \mathbb{R}^{n \times n}$ using $f = \Psi x$. For example, if f was an acoustic waveform and Ψ was the inverse FFT basis operator, x would be the Fourier coefficients of the waveform. In the traditional (lossy) source-channel sequential coding process, the largest m, where $m \ll n$, coefficients of x would be passed to the channel encoder. Let $y = \mathbf{R}x$ be the input to the channel coder, where $\mathbf{R} \in \mathbb{R}^{m \times n}$ is sub-matrix of $\mathbf{I_n}$ that defines the indices of x selected for transmission. The output at the sensor node would then be $z = \Omega y = \Omega \mathbf{R} \Psi^{-1} f$, where Ω is now of size $k \times m$. At the receiving end, the channel decoder first recovers $\hat{y}$ and thus $\hat{x}$ using (1) (replacing $\hat{f}$ with $\hat{y}$) and then $\hat{f} = \Psi \hat{x}$.

B. Compressive Sensing Fundamentals

The theory of compressive sensing asserts that the explicit compression step $x = \Psi^{-1} f$ does not need to be performed at the encoder and that a much smaller 'incoherent' transformation may be performed instead. We consider a sensing matrix $\Phi \in \mathbb{R}^{m \times n}$ that generates m $(m \ll n)$ of these incoherent measurements directly by projecting the signal f in its native domain[2] through $y = \Phi f$. In the usual synchronous sampling regime, Φ is an identity matrix I_m and $m = n$. When employing compressive sensing, however, the sensing matrix may be generated pseudo-randomly using various statistical distributions that ensure that sufficient unique information is captured with high probability. The questions that CS theory answers are: How can f be recovered from y? How many measurements m are required for accurate recovery and what sensing matrices Φ facilitate recovery? We summarize some key results from [4] and references therein.

The foundational argument behind compressive sensing is that although Φ is not full rank, x and hence f can be 'decoded' by exploiting the sparsity of x coupled with the sparsity promoting property of the ℓ_1 norm. To accomplish this, we view y as being generated through $y = \Phi \Psi x = \mathbf{A}x$ instead of through $y = \Phi f$. Now, while there are infinitely many solutions to $y = \mathbf{A}x$, the CS reconstruction procedure selects the one with the least sum of magnitudes by solving a constrained ℓ_1 minimization problem [7]:

$$\hat{x} = \arg\min_{\tilde{x}} \|\tilde{x}\|_{\ell_1} \ \text{ s.t. } \ y = \mathbf{A}\tilde{x}, \tag{2}$$

where, $\|x\|_{\ell 1} \triangleq \sum_{i=1}^{n} |x_i|$ and, f is recovered using $\hat{f} = \Psi \hat{x}$ as before. To guarantee that the solution from (2) is exact, a notion termed the restricted isometry property (RIP) was introduced. We will return to the RIP shortly, but first explain how the above procedure could be extended to handle missing data.

C. Handling Data Losses Compressively

Since the compressively sampled measurements in y are a compact representation of f, a valid scheme to protect y from channel erasures would be to feed it to a channel coding block as before. Thus, the sensor would now emit the coded measurements $z = \Omega y = \Omega \Phi f$, with Ω being of size $k \times m$ $(k > m)$. At the receiver, recovering f proceeds by first recovering $\hat{y}$ from $z' = \mathbf{C}z$ using (1) and then $\hat{x}$ using (2), if channel decoding succeeds.

If we consider each step in the above process, we see that compressive sampling concentrated the signal information in a set of m measurements and then channel coding dispersed that information to a larger set of k measurements. A natural question to ask is how the dispersion scheme differs in essence from the concentration scheme and whether they can be unified. The answer to this question is the crux of this paper.

We argue that compressive sensing not only concentrates but also spreads information across the m measurements acquired. This perspective is backed by Theorem **2** (below) and is the primary reason for the logarithmic rate overhead experienced by CS practitioners. Based on this observation, we propose that, an efficient strategy for improving the robustness of data transmissions is to augment the sensing matrix Φ with e additional rows generated in the same way as the first m rows. These extra rows constitute extra measurements, which, under channel erasures will ensure that sufficient information is available at the receiver. Note that oversampling in the native domain of f is also a valid strategy, but is highly inefficient. On the other hand, we will show next that if $k = m + e$ incoherent measurements are acquired through "compressive oversampling" and e erasures occur in the channel, the CS recovery performance will equal that of the original sensing matrix with a pristine channel with high probability (w.h.p.). We denote the augmented sensing matrix as $\bar{\Phi} \in \mathbb{R}^{k \times n}$ and the samples received at the decoder would be $z' = \mathbf{C}\bar{\Phi} f$. The decoding and recovery procedures for this case are now performed in one-step using (2), but constrained by z' (instead of y') to incorporate augmentation and losses: $z' = \mathbf{C}\bar{\Phi}\Psi x = \mathbf{C}\bar{\mathbf{A}}x = \bar{\mathbf{A}}' x$.

To understand intuitively why such an approach might work, assume that both Φ and $\bar{\Phi}$ are generated randomly with each element being an instance of an *i.i.d.* random variable. From our earlier discussion on viewing missing measurements as missing rows in the coding matrix, we observe that with $k = m + e$ and e missing measurements, $\bar{\Phi}' = \mathbf{C}\bar{\Phi}$ would be of size $m \times n$. Now, since each element of $\bar{\Phi}$ is *i.i.d.* and the erasure channel does not modify its value, we can view $\bar{\Phi}'$ as being generated with m rows to begin with, just like Φ. So, while Φ and $\bar{\Phi}'$ will not be identical, their CS reconstruction performance, which depends on their statistical properties and their size, will be equal (with high probability). We explain this analytically in the following section.

[2] The native domain for typical analog-to-digital conversion is time, but in some cases like photonic ADCs [3], sampling occurs in the frequency domain.

D. Robustness of CSEC to Erasures

We can show that CS oversampling is not only a valid erasure coding strategy, but also an efficient one. In particular, we would like to show that if we augment the sensing matrix to include e extra measurements and any e from the set of $k = m + e$ measurements are lost (randomly and independently) in transmission, the performance of CS reconstruction is equal to that of the original un-augmented sensing matrix (with high probability). To accomplish this, we rely on results from compressive sensing theory. We define the restricted isometry constant δ_s of a matrix $\mathbf{A} = \Phi\Psi$ and reproduce a fundamental result from [4] that links δ_s to CS performance.

Definition 1. [4] For each integer s=1,2,..., define the isometry constant δ_s of a matrix $\mathbf{A}$ as the smallest number such that:

$$(1-\delta_s)\|x\|_{\ell_2}^2 \le \|\mathbf{A}x\|_{\ell_2}^2 \le (1+\delta_s)\|x\|_{\ell_2}^2 \quad (3)$$

holds for all s-sparse vectors x. A vector is said to be s-sparse if it has at most s non-zero entries.

Theorem 2. [4] Assume that $\delta_{2s} < \sqrt{2} - 1$ for some matrix $\mathbf{A}$, then the solution $\hat{x}$ to (2) obeys:

$$\|\hat{x} - x\|_{\ell_1} \le C_0 \cdot \|x - x_s\|_{\ell_1} \quad (4)$$

$$\|\hat{x} - x\|_{\ell_2} \le \frac{C_0}{\sqrt{s}} \cdot \|x - x_s\|_{\ell_1} \quad (5)$$

for some small positive constant C_0 and x_s is an approximation of a non-sparse vector with only its s-largest entries. In particular, if x is s-sparse, the reconstruction is exact.

This theorem not only guarantees that CS reconstruction will be exact if $\delta_{2s}(\mathbf{A}) < \sqrt{2} - 1$ for an s-sparse signal, but also that if x is not s-sparse, but is compressible, with a power-law decay in its coefficient values, ℓ_1 minimization will result in the best s-sparse approximation of x, returning its largest s coefficients. We will return to this property when we compare the performance of CSEC with traditional erasure coding. Note also, that this is a deterministic result.

CS theory also suggests mechanisms to generate $\mathbf{A}$ matrices that satisfy the RIP with high probability. For example, it has been shown in [7] that the matrix $\mathbf{A}$ can be constructed randomly using i.i.d. random variables, with a Gaussian $A_{ij} = \mathcal{N}(0, 1/n)$ distribution or an equi-probable Bernoulli $A_{ij} = \pm 1/\sqrt{n}$ distribution. Using such matrices in low-power sensing devices, however, is difficult since implementing the sensing matrix $\Phi = \mathbf{A}\Psi^{-1}$ involves sampling and buffering f and computing $y = \Phi f$ explicitly through complex floating point operations. It was also shown in [7] and [16] that if $\mathbf{A}$ is constructed by randomly selecting rows of a Fourier basis matrix, the number of measurements obeys:

$$s \le C_2 \cdot \frac{m}{\log^4(n)} \quad (6)$$

with high probability. This is a significant result indeed because it implies that, if the signal is sparse in the Foureier domain, $\Phi = \mathbf{A}\Psi^{-1}$ is essentially an $m \times n$ random sampling matrix constructed by selecting m rows independently and uniformly from an identity matrix $\mathbf{I}_n$. This Φ is trivially implemented by pseudo-randomly sampling f, m times and communicating the stream of samples and their timestamps to the fusion center. Matrix Φ can then be recreated at the fusion center from the timestamps. The limitation on Ψ being the Fourier basis was removed in [16], which showed that the bound in (6) extends to any dense ortho-normal basis matrix Ψ with uniform random sampling.

Assume that the transmission channel can be modeled using an independent Bernoulli process with mean loss probability p. Thus, the likelihood of any measurement being dropped in this memoryless erasure channel is equal and is p. To show now that CSEC with random sampling is efficient for this channel model, we need to show that reconstruction performance with $\mathbf{A} = \Phi\Psi$, where $\Phi \in \mathbb{R}^{m\times n}$ and $\overline{\mathbf{A}'} = \mathbf{C}\bar{\Phi}\Psi$, where $\bar{\Phi} \in \mathbb{R}^{k\times n}$ is equal with high probability when $k = m/(1-p)$. The factor $1-p$ denotes the ratio of measurements lost in the channel. However, note that since Ψ is an ortho-normal basis matrix, it is equivalent to show that sensing performance with Φ and $\bar{\Phi}' = \mathbf{C}\bar{\Phi}$ is identical w.h.p.

Our approach considers the Fourier random sampling strategy, which constructs a Φ matrix by selecting samples from f at random. For the bound in (6) to hold for matrix Φ, two conditions need to be met:

a) At least m samples need to be selected
b) The indices should be selected using a uniform random distribution so that each sample is equi-probable.

To show that $\bar{\Phi}'$ results in identical performance (w.h.p.) to Φ, we need to show that the above conditions hold equally. We first show that $\bar{\Phi}'$ satisfies condition (b). When the channel is memoryless with a loss probability p, an average of $k \cdot p$ samples are lost in transmission and only k' samples are received. Let $\mathcal{S}_\Phi$ denote the set of sample indices that were selected using $y = \Phi f$ for the random sampling case. Therefore, the cardinality of $\mathcal{S}_\Phi$ would be $|\mathcal{S}_\Phi| = m$. Similarly, let the set of indices chosen in $\bar{\Phi}$ be labeled as $\mathcal{S}_{\bar{\Phi}}$ and the sample indices received by the decoder be $\mathcal{S}_{\mathbf{C}\bar{\Phi}}$, where $|\mathcal{S}_{\bar{\Phi}}| = k$ and $|\mathcal{S}_{\mathbf{C}\bar{\Phi}}| = k'$. Since Φ is constructed randomly and uniformly, the probability of a sample i being selected from f is:

$$\Pr\left[i \in \mathcal{S}_\Phi \mid |\mathcal{S}_\Phi| = m\right] = \frac{m}{n} \quad (7)$$

and for the oversampling case is:

$$\Pr\left[i \in \mathcal{S}_{\bar{\Phi}} \mid |\mathcal{S}_{\bar{\Phi}}| = k\right] = \frac{k}{n} \quad (8)$$

Claim 3. If we transmit k randomly chosen samples over an independent Bernoulli channel with a probability of lost transmission over the channel as p, the probability of the i^{th} sample of f being received in the k' samples is:

$$\Pr\left[i \in \mathcal{S}_{\mathbf{C}\bar{\Phi}} \mid |\mathcal{S}_{\mathbf{C}\bar{\Phi}}| = k'\right] = \frac{k'}{n} \quad (9)$$

Proof. This result is intuitive and straightforward to prove.

$$\Pr\left[i \in \mathcal{S}_{\mathbf{C}\bar{\Phi}} \mid |\mathcal{S}_{\mathbf{C}\bar{\Phi}}| = k'\right] = \frac{\Pr\left[i \in \mathcal{S}_{\mathbf{C}\bar{\Phi}}, |\mathcal{S}_{\mathbf{C}\bar{\Phi}}| = k'\right]}{Pr\left[|\mathcal{S}_{\mathbf{C}\bar{\Phi}}| = k'\right]}$$

$$= \frac{\Pr\left[\text{receiving sample } i \text{ correctly}\right] \cdot \Pr\left[|\mathcal{S}_{\mathbf{C}\bar{\Phi}}| = k' - 1\right]}{Pr\left[|\mathcal{S}_{\mathbf{C}\bar{\Phi}}| = k'\right]}$$

$$= \frac{(1-p) \cdot \Pr\left[\text{selecting sample } i\right] \cdot \Pr\left[|\mathcal{S}_{\mathbf{C}\bar{\Phi}}| = k' - 1\right]}{Pr\left[|\mathcal{S}_{\mathbf{C}\bar{\Phi}}| = k'\right]}$$

$$= \frac{(1-p)\cdot \Pr\left[i \in \mathcal{S}_{\bar{\Phi}} \mid \left|\mathcal{S}_{\bar{\Phi}}\right| = k\right]\cdot \binom{k-1}{k'-1}\cdot p^{k-k'}(1-p)^{k'-1}}{\binom{k}{k'} p^{k-k'}(1-p)^{k'}}$$

$$= \frac{(1-p)k/n}{k/k'}\cdot\frac{p^{k-k'}(1-p)^{k'-1}}{p^{k-k'}(1-p)^{k'}} = \frac{k'}{n}$$

This means that, if the channel is modeled as an independent Bernoulli process and the input sample distribution is equiprobable over n samples, the output index distribution is also equiprobable over the set of correctly received samples. This proves condition (b). If we increase the number of samples by the ratio lost in the channel such that $k = m/(1-p)$, then $\mathrm{E}[k'] = m$ and condition (a) is satisfied as well:

$$\Pr\left[i \in \mathcal{S}_{\mathbf{C}\bar{\Phi}} \mid \left|\mathcal{S}_{\mathbf{C}\bar{\Phi}}\right| = k'\right] = \Pr\left[i \in \mathcal{S}_{\Phi} \mid \left|\mathcal{S}_{\Phi}\right| = m\right] = \frac{m}{n} \qquad (10)$$

We, therefore, conclude that the sample indices in $\bar{\mathbf{\Phi}}'$ are statistically indistinguishable from the indices in $\mathbf{\Phi}$ for a memoryless channel and that the bound for the required number of measurements holds equally (with high probability).

E. CSEC Reconstruction when Redundancy is Insufficient

While the above result indicates that signal recovery using CSEC is exact if the redundancy $k-m$ is higher than the number of erasures $k-k'$, it can also be shown that when $k' < m$, recovery can still proceed but results in an approximation of the signal. This is in contrast to traditional erasure coding schemes that necessitate that the matrix $\mathbf{C\Omega}$ be invertible for any reconstruction to occur. To prove this we use Thm. **2** when applied to compressible signals. Assume that the signal of interest f in its compressed form x has its ordered coefficients decaying according to a power law such that $|x|_{(l)} \le C_r \cdot l^{-r}$, where $|x|_{(0)} \ge |x|_{(1)} \ge \cdots \ge |x|_{(n)}$ and $r \ge 1$.

Assume also that the bound in (6) is satisfied in equality for some choice of m and s. Now, when $k' < m$, k' does not meet the bound for sparsity s. However, k' is guaranteed to satisfy the bound for some lower sparsity $s' \le s$ (by extension from [16]). For the set of k' measurements received, Thm. **2** guarantees that CS reconstruction will result in the best s'-sparse approximation of x, returning its largest s' coefficients. This implies, then, that the increase in the ℓ_1 norm of the reconstruction error with k' measurements will be limited to $\epsilon \le C_0 \left\|x_s - x_{s'}\right\|_{\ell_1}$ with high probability. We empirically study the effect of this reconstruction error on the probability of recovery in Sec. III.B.

III. Evaluation Results

In order to verify the performance of compressive erasure coding, we analyze the sampling matrix $\bar{\mathbf{A}}'$ that identifies which measurements were received at the decoder.

A. Verifiable Conditions using RIP

1) For a Memoryless Erasure Channel

We model the erasure introduced by the transmission channel with an average measurement loss probability, p. We initially assume an independent Bernoulli process. The question we would like to address is how the loss of $k\cdot p$ measurements dropped (on the average) affects CS reconstruction. From Thm. **2**, we understand that reconstruction accuracy depends on the RIP constant δ_{2s} of $\mathbf{A}$. To evaluate the extent of performance loss through the erasure channel, we can thus rely upon quantifying $\delta_s(\mathbf{A})$. Computing $\delta_s(\mathbf{A})$ exactly from Def. **1**, however, is exhaustive because it is defined over all s-sparse vectors. We approximate it by evaluating the eigenvalues of its Grammian [2] over 10^3 random $s\times n$ sub-matrices. Increasing this number to 10^6 results in little improvement.

We first generate a random sampling matrix $\mathbf{A} = \mathbf{\Phi\Psi}$ of size $m\times n$ as described in Sec. II.B. This sensing matrix is modified by the channel so that $\mathbf{A}' = \mathbf{CA}$. We also have an augmented sensing matrix generated at the source $\bar{\mathbf{A}} = \bar{\mathbf{\Phi}}\mathbf{\Psi}$ of size $k\times n$ with $k > m$ and its received counterpart $\bar{\mathbf{A}}' = \mathbf{C}\bar{\mathbf{A}}$. Testing the performance of CSEC numerically then proceeds by comparing whether $\delta_s(\bar{\mathbf{A}}') \le \delta_s(\mathbf{A})$. Equality ensures that the CS decoder would be able to achieve reconstruction accuracy identical to an un-augmented sensing matrix with a pristine channel. If $\delta_s(\bar{\mathbf{A}}') < \delta_s(\mathbf{A})$, it means that the decoder has more measurements through $\bar{\mathbf{A}}'$ than through the original $\mathbf{A}$ and would lead to a higher probability of exact recovery.

The result from this calculation for a Monte Carlo simulation over 1000 256×1024 random sampling matrices with the Fourier basis for reconstruction is shown in Fig. 3. The dotted blue curve labeled "No Loss" indicates $\delta_s(\mathbf{A})$ forming the baseline for our comparison. The shading illustrates the min-max values over all choices of $\mathbf{\Phi}$. With loss probability $p = 0.2$, we see a mean increase in RIP constant, which implies that the sparsity for guaranteed ℓ_1 reconstruction drops. In this case, we see that the sparsity drops from about 6 to 4 based on the $\delta_{2s} < \sqrt{2} - 1$ bound (gray horizontal line) from Thm. **2**. Note, though, that while this bound is known to be conservative, enumerating the RIP constant in this way clearly indicates the loss in reconstruction accuracy that may be expected by losing 20% of the sampled measurements.

From Clm. **3**, we see that the probability distribution of indices extracted from $\mathbf{\Phi}$ and $\mathbf{\Phi}' = \mathbf{C\Phi}$ are identical when $\mathbf{C}$ comes from a memoryless (independent Bernoulli) channel. This means that, if the channel is not congested, increasing the sensing rate by a factor of $p/(1-p)$ will restore the delivery rate to $k' = m$ on average. The effect of this increase is substantiated in Fig. 3 and establishes $\delta_s(\bar{\mathbf{A}}') \approx \delta_s(\mathbf{A})$ for the independent Bernoulli channel. We see that not only does the mean RIP constant improve to its original value but that the range of variation also recovers to the "No Loss" baseline. Note also, that the minimum values of $\delta_s(\bar{\mathbf{A}}')$ are below $\delta_s(\mathbf{A})$ suggesting that some instances of $\bar{\mathbf{\Phi}}$ coupled with channel loss actually deliver better-than-baseline performance.

To compare performance across different sampling schemes, we further evaluate the RIP constant for a sensing matrix that is constructed using the Gaussian random projection method described earlier. In this case, the reconstruction is performed in the identity domain with $\mathbf{A} = \mathbf{\Phi}$. It has been shown in [7] that the Gaussian projection technique has equivalent performance across any ortho-normal reconstruction basis and the identity matrix was chosen for computational ease. The result of this computation is shown in Fig. 4 for a memoryless lossy channel with $p = 0.2$. Here too, we observe that while the RIP constant is higher in the lossy case, increasing the rate by

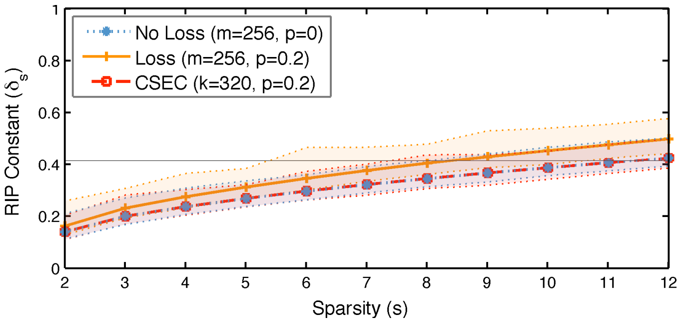

Figure 3. RIP constant for random sampling with a memoryless channel.

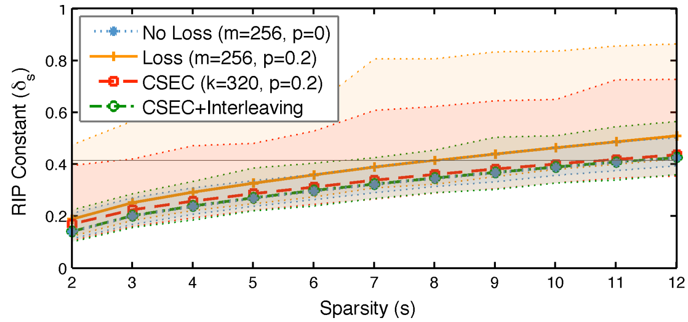

Figure 6. RIP constant for random sampling with a Gilbert-Elliott channel.

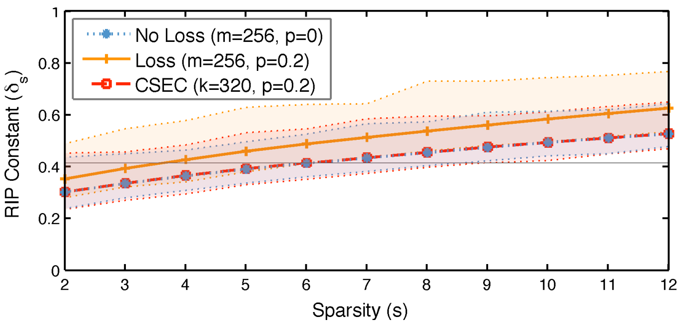

Figure 4. RIP constant for Gaussian projections with a memoryless channel.

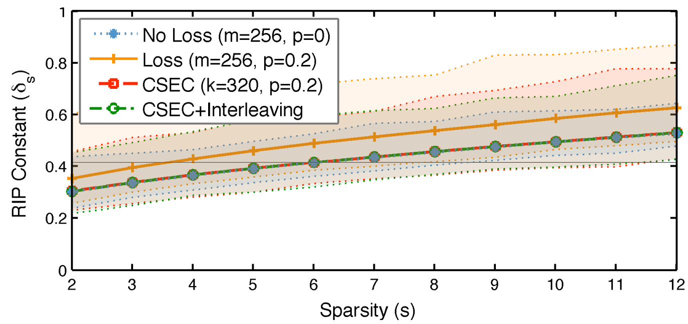

Figure 5. RIP constant for Gaussian projections with a Gilbert-Elliott channel.

the amount lost in the channel recovers the performance guaranteed by compressive sensing.

2) Interleaving for Bursty Channels

Realistic wireless channels exhibit bursty behavior [17]. To estimate the effect of CSEC performance with bursty channels, we use the popular Gilbert-Elliott (GE) model [10], which is both tractable to use and accurate in describing many wireless channels (including those in mobile environments [18]). GE channels are modeled using a stationary discrete-time binary Markov process. Within each state, marked good and bad, the probability of loss is assumed to be $p_g = 0$ and $p_b = 1$ respectively. The probability of transition from one state to another is marked as p_{gb} and p_{bg}. To maintain consistency with the memoryless channel performance studies, we compute transition probabilities from average loss probability p and expected burst size b using standard relationships: $p_{bg} = 1/b$ and $p_{gb} = p/(b(1-p))$.

Fig. 6 shows the RIP constant δ_s at loss probability at $p = 0.2$ and with an expected loss burst of $b = 8$ samples. While $b = 8$ constitutes an extreme condition of burstiness, it is instructive to see its effects on CS recovery. Thus, the same number of samples is delivered to the fusion center as with the memoryless channel on average, but with a modified index distribution. The effect of this change is immediately evident in the variation of δ_s in Fig. 6 indicating that some sensing matrices are particularly bad for a GE channel. While an increase in sensing rate improves the mean RIP constant (though not reaching the baseline), the variance remains quite high.

The variance issue can be resolved by applying randomized interleaving prior to transmission, which results in a roughly uniform distribution of the sample losses [13]. It can be shown that interleaving recovers the original index distribution (up to a bound) for random sampling and the green dotted curve (coincident with the baseline) in Fig. 6 illustrates this empirically. Note, however, that interleaving requires buffering y (though not f), which increases decoding latency. Interestingly, we observe that the Gaussian random projection technique in Fig. 5 is unaffected by interleaving. In fact, using Gaussian projections delivers near baseline performance with or without interleaving (min-max variation is not perfect). This is because the sensing matrix Φ is dense (compared to the one for random sampling) with each element within it being *i.i.d* Gaussian. This means that every measurement in y has a random, independent but statistically identical contribution from every element in f. Since interleaving the measurements y is equivalent to shuffling the rows of the matrix Φ, interleaving does not affect the statistical properties of Φ.

A more extensive treatment of this "democratic" property of Gaussian sensing matrices can be found in a recent report by Davenport, et al. [24], which analyses the performance from a theoretical perspective. The democracy argument has been employed in [25] for the novel application of handling saturation errors in analog-to-digital quantizers, in a fashion similar to what we propose in CSEC.

B. *Signal Reconstruction Performance*

Evaluating the RIP constant provides theoretical insight into what the performance gain would be when using CSEC. In this section, we study the practical implications by evaluating the probability P_{ex} with which CSEC could deliver the original signal exactly. We do this by performing a Monte-Carlo simulation over 10^4 random instances of a length 256 sparse signal and computing how often CS erasure coding results in exact recovery. Figs. 7 to 11 illustrate the comparative performance of using Fourier random sampling (left) and the Gaussian projection method (right) for CSEC. For each plot, the "No Loss" curve indicates the baseline with $m = 64$ (4:1 compression) and

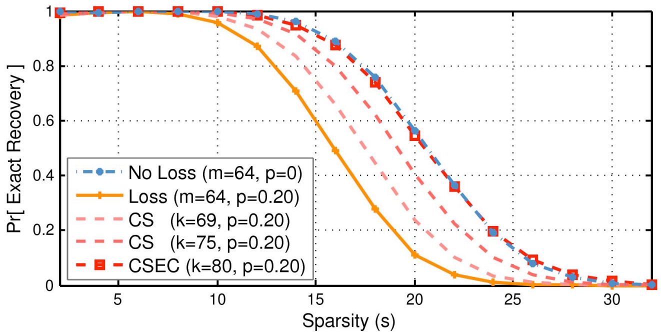

Figure 7. Probablity of recovery for random sampling with a memoryless erasure channel.

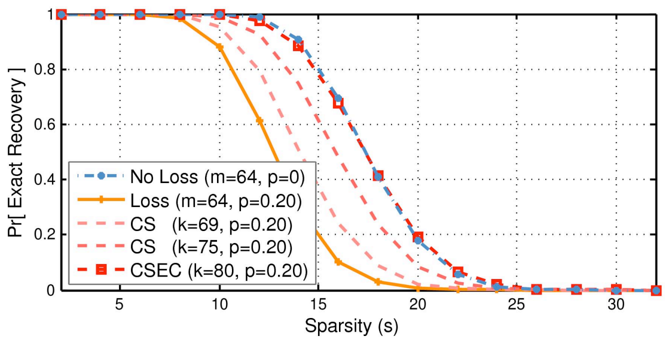

Figure 8. Probablity of recovery for Gaussian projections with a memoryless erasure channel.

the "Loss" curve indicates the probability when no over-sampling is performed. The "CSEC" (red) curve indicates P_{ex} with compressive oversampling at $k = m/(1-p)$ and the two "CS" curves indicate intermediate values with $m < k < m/(1-p)$. The x-axis indicates the number of non-zero coefficients in x.

Three channel models have been used to generate these figures. Figs. 7 and 8 mimic the channel model used in Figs. 3 and 4, a memoryless erasure channel modeled as an independent Bernoulli process with $p = 0.2$. We see for both Fourier random sampling and Gaussian projections that, when $k - k' = k \cdot p = 16$ measurements are lost on the average, $k = m/(1-p) = 80$ recovers performance to the original $m = 64$ level. Observe that if the bound in (6) is not met (beyond about $s = 10$), the performance for a particular k drops gradually with s. Note, also, that P_{ex} decays quicker to 0 in the case of Gaussian projections and, we see while comparing with Figs. 3 and 4, this is because the RIP constant is also higher for the latter. The intermediate values of k ($k = 69$ and $k = 75$) in both cases deliver intermediate levels of quality as predicted in Sec. II.E.

Figs. 9 and 10 use the same Gilbert-Elliott channel model as Figs. 6 and 5 with $p = 0.2$ and $b = 8$. It is striking to note that due to the burstiness of the channel, the performance of neither Fourier random sampling nor Gaussian projections reaches the baseline for low sparsity levels. Further, while the highest s for which $P_{ex} = 1$ has gone down substantially for the lossy scenarios, the slope of the P_{ex} curve is also reduced. The reason for this is that the distribution of received sample lengths, k', across the Monte-Carlo runs is skewed and asymmetric about the mean for bursty channels, whereas it is symmetric about $k(1-p)$ and is Gaussian for a memoryless channel. As an ex-

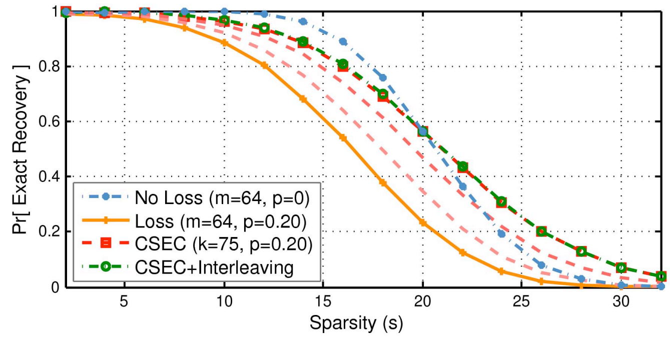

Figure 9. Probablity of recovery for random sampling with a Gilbert-Elliott channel model.

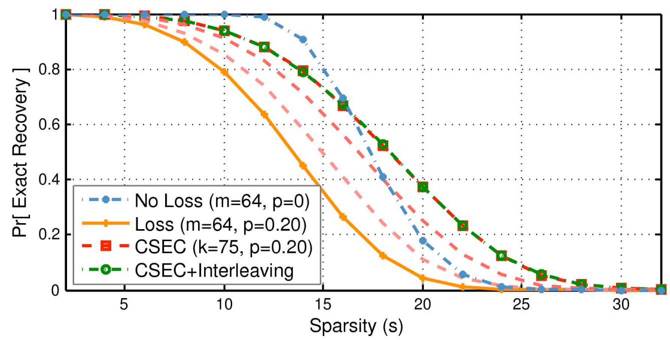

Figure 10. Probablity of recovery for Gaussian projections with a Gilbert-Elliott channel model.

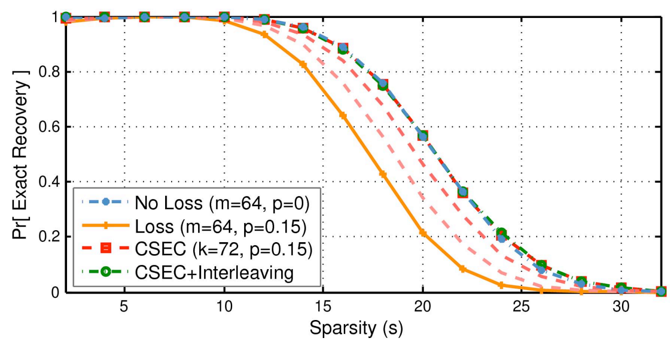

Figure 11. Probablity of recovery for random sampling with a real 802.15.4 channel using GE model.

ample, the mean value of k' for CSEC across runs is ${}^{k'} \approx 64$ for both the Bernoulli channel and the GE channel but their standard deviations are $\sigma^{k'}{}_{Bern} \approx 3$ and $\sigma^{k'}{}_{GE} \approx 10$ respectively. It's interesting, though; the sample length distribution is skewed toward higher k' and the result is that the probability of reconstruction at larger sparsity levels is actually higher than the baseline. An unexpected result from Fig. 9 is that interleaving makes little or no difference to Fourier random sampling.

Fig. 11 has been generated using a wireless network trace from the CRAWDAD database [11], which provides extensive network performance datasets collected from a wide array of conditions. The particular trace we selected used sensor nodes with an IEEE 802.15.4 radio transceiver placed about 12m apart between two different floors of a university building. This trace had the highest loss probability and burstiness across the 27 traces collected with $p \approx 0.15$ and $b = 1.2$. We built a GE channel model based off the trace and simulated the probability of exact recovery as before. There is very little burstiness in the

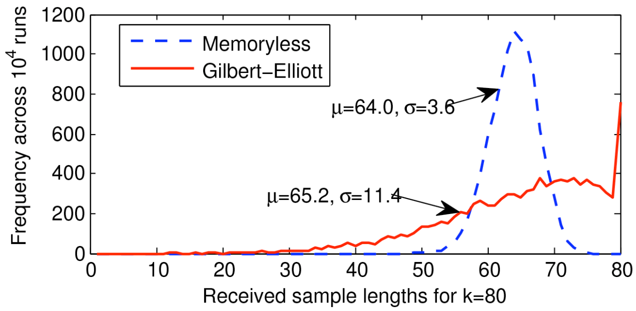

Figure 12. Comparison between received sample length distributions resulting from transmission through memoryless and Gilbert-Elliott channels.

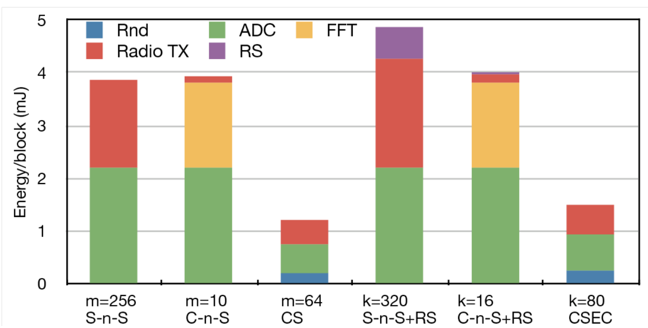

Figure 13. Energy consumption comparison for different sampling strategies (form left to right: Sample-and-Send (S-n-S), Compress-and-Send (C-n-S), Compressive Sensing (CS), S-n-S with Reed-Solomon encoding, C-n-S with Reed-Solomon encoding and CS Erasure Coding by Oversampling).

channel and Fig. 11 shows that CSEC will be able to deliver near baseline performance with either Fourier random sampling or Gaussian projections (the plot for Gaussian projections was omitted since it was identical to Fig. 8).

C. *CSEC Implementation Costs*

We can quantify the energy efficiency gains that CS promises too. In particular, we use random sampling with $n = 256$ and compare it to two cases – first, where a standard $(255, 223)$ Reed-Solomon (RS) [15] code, a popular BCH code, is applied to a set of 256 raw 16-bit samples and second, where RS is applied to a compressed version of the signal. We assume the signal is sparse ($s \leq 10$) in the Fourier domain and use 256-point FFT for source compression. Fig. 13 shows this comparison, which also includes energy consumption costs without RS. The data has been extracted using a cycle and energy accurate instruction-level simulator [21] available for the popular MicaZ sensor platform. While the analysis is specific to this platform, the insight from these results can be applied more generally.

We have split the costs among five blocks, which are significant for the comparison – random number generator (for CS), ADC, FFT processing, radio transmission and Reed-Solomon coding. The total energy consumption of sample-and-send and compress-and-send is almost equal without RS, with the radio taking a large chunk of the former and the FFT routine consuming half of the latter. Notice also, that the ADC energy consumption is substantial since both these techniques need to operate on the entire signal vector. On the other hand, the CS routine at $m = 64$ (4:1 compression) requires a fraction of the ADC and radio, but incurs an overhead for generating random numbers. The current implementation uses an inexpensive 16-bit LFSR for pseudo-random number generation.

When the data is RS encoded before transmission, the energy consumption of the sample-and-send strategy jumps considerably, whereas the increase for compress-and-send is negligible, because it is sending at most 10 coded symbols (with 6 parity symbols). We chose $s \leq 10$ since that is the threshold below which $m = 64$ in the lossless case and $k = 80$ for a memoryless erasure channel result in exact CS recovery (refer Figs. 7 and 11). This means that, with a $p = 0.2$ memoryless channel, all three strategies on the right will deliver equivalent recovery performance. When comparing encoding cost, however, CS erasure coding is 2.5× better than performing local source compression and 3× better than sending raw samples.

IV. Related Work and Discussion

Recovering from erroneous and missing data in communication systems employ ARQ retransmissions and forward error correction (FEC) routinely, sometimes simultaneously, at different layers of the communication protocol stack. For sensor networks, however, the simplicity of ARQ has retained it as the dominant form of error recovery. Many researchers have questioned this recently and evaluated FEC techniques through lab experiments.

For example, Schmidt, et. al. [20] focused on convolutional coding to show that a modified Turbo code is quite feasible on a MicaZ platform. While they report that the energy consumption using Turbo codes is about 3.5× of its un-coded counterpart, the overall energy efficiency is better considering retransmission costs, especially on high loss links. They also show that the computational complexity of Turbo encoding is practical, but only with low-rate data transfers (they tested with one packet every second). Jeong, et. al [19] proposed using a simpler error correcting code for only single and double-bit errors to reduce this computation burden. They illustrate, through experimental data that long error bursts are rare in static sensor networks and argue that the complexity of RS or LT codes is unwarranted. They show that their error correcting code reduces the packet drop rate almost to zero for outdoor deployments. However, due to a higher frequency of multiple-bit errors indoors, recovery remains imperfect there. We've shown that CSEC can provide both computational benefits and recovery performance that parallels state-of-the-art erasure correcting codes. To use CSEC, however, one must have a good understanding of the physical phenomena being acquired and the domain it can be compressed in.

Further, Wood, et. al. [22] recently reported the use of online codes in low-power sensor networks. Online codes are a form of digital fountain codes, such as the LT codes [12], but are simpler to encode and decode. They propose a lightweight feedback mechanism that allows the encoder to cope with variations in link quality rapidly and efficiently. They point out, however, that multiple parameters need to be tuned in order for the coding to be efficient. This is similar to the sensitivity of LT codes to the degree distribution [12]. While our current work has focused on block wise decoding and makes analogies to other linear block coding strategies such as BCH codes, CSEC can be used in a "rateless" mode as well, similar to fountain codes. Asif, et al. [23] demonstrate a way of streaming incoherent measurements and describe a homotopy based approach that performs iterative decoding as measurements are received. Using this technique, while exploiting the democracy property of sensing matrices [24], a stream of measurements

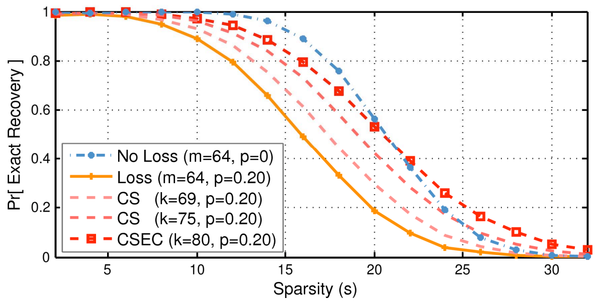

Figure 14. Probablity of recovery for random sampling with a memoryless channel and 8 measurements per packet.

that is transmitted through an erasure channel can be progressively decoded as measurements are gradually received.

We have described CSEC for handling erasures in a channel, but CSEC can be extended to correct for errors in the sensor transduction process too. This means that a controlled amount of sensor noise can be cleaned from the acquired measurements during the decompression process. This is achieved by using Basis Pursuit De-noising [8], which changes the equality constraint in (2) to an inequality to account for variations due to noise. Note, however, that since CSEC utilizes features of the physical phenomenon and operates on the acquired signal, and not on the modulated symbols transmitted through the wireless channel, CSEC is not useful for correcting symbol errors at a communication receiver. A better approach to tackling the latter using ℓ_1 minimization techniques is discussed by Candes and Tao in [6].

In Sec. III.A, we used the RIP constant of the sensing matrix as way of verifying its reconstruction performance. Another technique that was recently proposed, namely the null-space property [9], could also have been used. Until much recently, however, the null-space property was as difficult to compute as the RIP constant. In the future, we will not only use the null-space property for evaluating sensing matrices numerically but also study its use to analyze the recovery properties of CSEC, especially with Gaussian projections.

And finally, we note that the evaluation studies in Sec. III assumed that measurements are streamed to the receiver as they are acquired. If one packetizes the measurements for transmission, in a memoryless channel, the sample losses will no longer be independent and instead show high burstiness. An example of this is shown in Fig. 14, which shows the probability of recovery when 8 measurements are transmitted in every packet. We defer a detailed study of the effects of packetization in realistic wireless channels for future work.

V. Conclusion

We have explored the application of Compressive Sensing to handling data loss from erasure channels by viewing it as a low encoding-cost, proactive, erasure correction scheme. We showed that CS erasure coding is efficient when the channel is memoryless and employed the RIP to illustrate, that even extreme stochasticity in losses can be handled cheaply and effectively. We showed that for the Fourier random sampling scheme, oversampling is much less expensive than competing erasure coding methods and performs just as well. This makes it an attractive choice for low-power embedded sensing where forward erasure correction is needed.

References

[1] I. Akyildiz, D. Pompili, and T. Melodia, “ Underwater acoustic sensor networks: research challenges,” *Ad Hoc Networks*, 2005.

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constructive Approximation*, 2008.

[3] A. S. Bhushan, F. Coppinger, and B. Jalali, “Time-stretched analogue-to-digital conversion," *Electronics Letters*, 1998.

[4] E. Candès, “The restricted isometry property and its implications for compressed sensing,” *Comptes rendus-Mathématique*, 2008.

[5] E. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, 2006.

[6] E. Candes and T. Tao, “Decoding by linear programming,” *IEEE Transactions on Information Theory*, 2005.

[7] E. Candes and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?,” *Transactions on Information Theory*, 2006.

[8] S. Chen, D. Donoho, and M. Saunders, “Atomic Decomposition by Basis Pursuit,” *SIAM Journal on Scienti• c Computing*, 1998.

[9] A. Cohen, W. Dahmen, and R. DeVore, “Compressed sensing and best k-term approximation,” *American Mathematical Society*, 2009.

[10] E. Elliott, “ Estimates of error rates for codes on burst-noise channels,” *Bell Systems Technical Journal*, 1963.

[11] A. Iqbal, K. Shahzad, S. A. Khayam, and Y. Cho, “CRAWDAD trace niit/bit_errors/802.15.4/Traces_802.15.4.”, *CRAWDAD*, July 2008.

[12] M. Luby, “LT codes,” *IEEE Symposium on Foundations of Computer Science,* 2002.

[13] M. Mushkin and I. Bar-David, “Capacity and coding for the Gilbert-Elliot channels,” *IEEE Transactions on Information Theory*, 1989.

[14] R. Newton, S. Toledo, L. Girod, H. Balakrishnan, and S. Madden, “Wishbone: Pro• le-based Partitioning for Sensornet Applications,” *Networked Systems Design and Implementation (NSDI)*, 2009.

[15] I. Reed and G. Solomon, “Polynomial codes over certain • nite • elds,” *Journal of the Society for Industrial and Applied Mathematics*, 1960.

[16] M. Rudelson and R. Vershynin, “Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements,” *Information Sciences and Systems*, 2006.

[17] K. Srinivasan, M. Kazandjieva, S. Agarwal, and P. Levis, “The -factor: measuring wireless link burstiness,” in *ACM Conference on Embedded Networked Sensor Systems (SenSys)*, 2008.

[18] H. Wang and N. Moayeri, “Finite state Markov channel-a useful model for radio communications systems,” *IEEE Transactions on Vehicular Technology*, 1995.

[19] J. Jeong and C. Ee, “Forward error correction in sensor networks,” *International Workshop on Wireless Sensor Networks (WWSN)*, 2007.

[20] D. Schmidt, M. Berning, and N. Wehn, “Error Correction in Single-Hop Wireless Sensor Networks-A Case Study,” *Design, Automation, and Test in Europe (DATE)*, 2009.

[21] B. Titzer, D. Lee, and J. Palsberg, “Avrora: Scalable sensor network simulation with precise timing,” *IPSN*, 2005.

[22] A. D. Wood and J. A. Stankovic, “Online Coding for Reliable Data Transfer in Lossy Wireless Sensor Networks,” *Distributed Computing in Sensor Systems (DCOSS)*, 2009.

[23] M. Asif and J. Romberg, “Streaming Measurements in Compressive Sensing: ℓ_1 Filtering,” *Asilomar Conference on Signals, Systems and Computers*, 2008.

[24] M. Davenport, J. Laska, P. Boufounos, and R. Baraniuk, “A simple proof that random matrices are democratic,” *Rice University ECE Department Technical Report*, TREE-0906, Nov. 2009.

[25] J. Laska, P. Boufounos, M. Davenport, and R. Baraniuk, “Democracy in action: Quantization, saturation, and compressive sensing,” *Preprint.*